THE LANCET
Infectious Diseases

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: McClain MT, Constantine FJ, Nicholson BP, et al. A blood-based host gene expression assay for early detection of respiratory viral infection: an index-cluster prospective cohort study. *Lancet Infect Dis* 2020; published online Sept 24. https://doi.org/10.1016/S1473-3099(20)30486-2.

Supplementary Online Material

Page:

Supplemental Methods

Enrollment, Sampling, and Case Definitions:

Demographic information, baseline biospecimens and a nasopharyngeal swab were collected at initial enrollment. Confirmed Index Cases were sampled for PAXGene RNA tubes, serum, plasma, and nasal swabs at the time of illness identification for phenotyping, gene expression analyses and other future work. Viral etiology was assessed from nasal swabs using a commercial multiplex PCR assay (QIAGEN ResPlex II v2.0, Luminex RVP, or Biofire FilmArray RP). Healthy CCs living in proximity to an IC provided blood (PAXGene, serum, plasma) and respiratory samples daily for up to five consecutive days and were monitored for symptomatic conversion and viral shedding. ICs identified on average 5.8 CCs (range 1-11) to be serially monitored. Quantitative measurement of eight symptoms were recorded daily via electronic survey (cough, fever, headache, malaise, nasal congestion, nasal discharge, sneezing, and sore throat, each with severity 0-4 where 0=not present, 1=Mild, 2=Moderate, 3=Severe, 4=Very Severe). For analysis, an IC was identified as a subject with a cumulative symptom score of ≥6 on a single day. A CC was considered to be a symptomatic conversion if cumulative symptom score reaches ≥6 on a single day in the 5-day window. Symptom score cutoffs were derived from prior viral challenge studies, where a cutoff of a symptom sum of 6 for a single day was determined to provide the optimal differentiation between infected and uninfected subjects[9,17,32]. If a Close Contact had overlapping contact with two Index Cases who became symptomatic within 1 day of each other, then that Close Contact would be considered a CC for both Index Cases. Each subject being monitored in the study could become an IC at any time except during the observation window when they were a CC. 10 days after being named a CC for IC1, the same CC could then be monitored/sampled again as a CC for an unrelated IC2 if otherwise indicated by the protocol. Healthy controls (asymptomatic non-shedders) are defined as those with a modified Jackson score of 0 and a negative etiology test. For each longitudinal profile we denote time 'T' as the day corresponding to the maximum cumulative daily modified Jackson score for that individual subject.

Microbiologic evaluation

Due to variable production/availability of tests across study years, 3 different respiratory viral assays were used over the course of the study. Nasopharyngeal swabs were first tested via the ResPlex II v .0 viral \multiplex assay

(Qiagen) which detects influenza A and B, adenovirus B and E, parainfluenza virus 1 to 4, respiratory syncytial virus (RSV) A and B, human metapneumovirus (hMPV), human rhinovirus (HRV), coronavirus (229E, OC43, NL63, and HKU1), coxsackie/echo virus, and bocavirus, the Luminex xTag Fast, Respiratory Viral Panel, which includes influenza A, influenza A/H1, influenza A/H3, RSV A and B, coronavirus (229E, HKU1, NL63, and OC43), parainfluenza virus 1-4, hMPV, adenovirus, and bocavirus, or by the BioFire Respiratory Panel which includes testing for adenovirus, coronavirus (229E, OC43, NL63, and HKU1), hMPV, HRVs/Enterovirus, influenza A, influenza A/H1, influenza A/H3, influenza A/H1-2009, influenza B, parainfluenza virus 1-4, RSV, *Bordetella pertussis*, *Chlamydophila pneumoniae*, and *Mycoplasma pneumoniae*. Kits were prepared and run per manufacturer's protocols as previously described[8]. In order to standardize results across platforms, all Rhinovirus or Enterovirus positive results were combined into a single category, "Rhinovirus/Enterovirus".

Real-time PCR

Peripheral blood was collected in PAXgene™ Blood RNA tubes (PreAnalytiX), and total RNA extracted using the PAXgene™ Blood miRNA Kit (QIAGEN) employing the manufacturer's recommended protocol. RNA quantity and quality were assessed using Nanodrop 2000 spectrophotometer (Thermo-Fisher) and Bioanalyzer 2100 with RNA 6000 Nano Chips (Agilent). cDNA synthesis was performed using SuperScript VILO™ Master Mix (Invitrogen) according to manufacturer's instructions. Real-time PCR was performed using custom TLDA 384-well microfluidic cards with TaqMan Gene Expression Master Mix, and run on a ViiA7 Real-Time PCR System (Applied Biosystems). RT-PCR data was generated for all IC participants, for CCs developing symptoms (symptom score increased by ≥6) across their 5 day window, as well as matched healthy controls. Raw quantification cycle (Cq)values were exported for statistical analysis. The TaqMan assays for selected transcripts were chosen based on a tiered list of criteria that involved primers and probe location within a target, probe exon spanning behavior, potential off-target amplification, and assay availability. To maximize performance of the RT-PCR platform, we performed iterative re-derivation of the regression coefficients for the gene transcripts in the model across training and test splits, including model parameter selection via Grid Search. This yields a comprehensive look at the quality of prediction expected. When comparing infected individuals at the time of maximal symptoms against a random selection of uninfected subject time points (on each iteration) the model results in an auROC mean of 0·9(IQR 0·87-0·93). For the development of all the classifiers described in this work, the data were normalized to the PCR values

of the control genes such that data from a given individual or a single individual could be classified. Gene components of the TLDA signature (and controls genes) are listed in Table S2.

Supplemental Description of Statistical Analyses

Definitions and data preparation

RT-PCR data were obtained for a total of 36 host gene expression targets in 811 samples. Reference transcripts were used to internally normalize expression values from the signature. Specifically, the reference genes (FPGS, TRAP1, or DECR1) were selected due to their low coefficients of variation across all samples. The quantification cycle, or Cq, values were transformed into normalized relative quantities using the reference assays as previously described[9]. Cases were symptomatic shedders as defined by a modified Jackson score of 6 or greater along with a positive viral etiology test. The healthy controls were asymptomatic non-shedders, defined by a modified Jackson score of 0 and a negative etiology test. For the purposes of analyzing accuracy of the RT-PCR assay across all viral infections, all individuals with multiple positive PCR results were considered similarly to individuals with only one virus detected, as they likely represented a true viral infection regardless of whether the source was mono or polymicrobial. However, for the purpose of determining virus-specific performance metrics (for instance in Figure 3), only subjects with a single virus detected during the episode were considered."

Training and Test Data Overview

We fit a statistical model to predict the probability that a subject is (or will become) a symptomatic shedder. To do so, we randomly assigned subjects in a 1:1 ratio to either the training dataset or test dataset. We then trained on all available timepoints within the training dataset by randomly selecting a single timepoint per subject ('T-3', 'T-2', 'T-1', 'T', 'T1', or 'IC'). This process is repeated for each iteration utilized during model development, as described below. This approach leads to a greater variation in the training data, as we have data from subjects before and after peak symptoms have appeared: thus the model sees the entire course of illness. For the test dataset, we considered all of time points individually.

Statistical Model

We train and update an L2-penalized logistic regression model in Tensorflow[34]. Given a subject's (normalized) gene values as an input, the model's output is a probability that this subject is a symptomatic shedder. As there are 36 genes, we have 36 coefficients, plus a bias/intercept that are learned during training. Additionally, we trained models to obtain different sets of regularization parameters. Overall, 25 instances of the model were trained where each instance has a different randomly generated train/test split. This is done to average out any splitting effects. Then, for a single model instance, we use 50 training iterations. In each iteration, we update the timepoints used for training: i.e., for each subject in the training set for each iteration, we select a timepoint at random. The model therefore sees a more complete representation of the subjects' time course both within and across iterations. In each iteration, we use the previous iteration's model parameters (logistic regression coefficients) as a starting point, and then train and update the parameters using the new training data. After all training was completed, we stored the result from the last iteration of each model instance. For testing, all of the subjects' timepoints were used independently. To select the regularization parameter for each of the 25 model instances, we trained with several different values and selected the value of the parameter that led to the lowest coefficient of variation from the area under the receiver operating characteristic (auROC) statistics across all test timepoints.

The model's classification accuracy was assessed by computing the auROC on the independent test participants across viruses and at each relevant timepoint. We then calculated the true positive rate (TPR) and false positive rate (FPR), which required thresholding of output (predicted) probabilities. This was selected by fixing a desired FPR (e.g., 0.2) and calculating the associated TPR at all timepoints [Figure s5]. We used a similar approach to define performance characteristics for identifying cases vs. controls using individual genes in the signature as well as symptom scores.

Table s1. Demographics

Subject Demographics	
Gender:	
Male	708 (48%)
Female	757 (52%)
Age (range):	
	18.2 (18-25)
Race, n (%)	
Native American	17 (1%)
Asian	385 (26%)
African American	133 (9%)
White	841 (57%)
Other	89 (6%)

Table s2. Student Cohort enrollment

	Total Monitored	Index Cases	Close Contacts
2009-2010	448	78	126
2010-2011	404	86	122
2011-2012	420	50	169
2013-2014*	88	15	56
2014-2015	105	35	82
Total	1465	264	555

* Truncated season, Spring Enrollment Only

Table s3. Viral Co-infections

Viruses	n
Coxsackie/Echo and HRV/Entero	16
Parainfluenza and RSV	1
Coronavirus and RSV	1
Coronavirus and HRV	1
Coronavirus and Coxsackie/Echo	1
Adeno and hMPV	1
Coxsackie/Echo and Parainfluenza	1

Table s4. Genes included in the signature

Gene ID	Gene Annotation
CCL2	C-C motif chemokine ligand 2(CCL2)
RTP4	receptor transporter protein 4(RTP4)
OAS3	2'-5'-oligoadenylate synthetase 3(OAS3)
RSAD2	radical S-adenosyl methionine domain containing 2(RSAD2)
IFI44L	interferon induced protein 44 like(IFI44L)
OAS1	2'-5'-oligoadenylate synthetase 1(OAS1)
OAS2	2'-5'-oligoadenylate synthetase 2(OAS2)
LY6E	lymphocyte antigen 6 complex, locus E(LY6E)
ISG15	ISG15 ubiquitin-like modifier(ISG15)
XAF1	XIAP associated factor 1(XAF1)
MX1	MX dynamin like GTPase 1(MX1)
HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5(HERC5)
SERPING1	serpin family G member 1(SERPING1)
IFI44	interferon induced protein 44(IFI44)
IFIT3	interferon induced protein with tetratricopeptide repeats 3(IFIT3)
TNFAIP6	TNF alpha induced protein 6(TNFAIP6)
IFIT2	interferon induced protein with tetratricopeptide repeats 2(IFIT2)
IFI27	interferon alpha inducible protein 27(IFI27)
OASL	2'-5'-oligoadenylate synthetase like(OASL)
IFIT1	interferon induced protein with tetratricopeptide repeats 1(IFIT1)
ATF3	activating transcription factor 3(ATF3)
LAMP3	lysosomal associated membrane protein 3(LAMP3)
IFIT5	interferon induced protein with tetratricopeptide repeats 5(IFIT5)
FPGS	folylpolyglutamate synthase(FPGS)
IFI6	interferon alpha inducible protein 6(IFI6)
SEPT4	Septin 4 (SEPT4)
SIGLEC1	sialic acid binding Ig like lectin 1 (SIGLEC1)
TRAP1	TNF receptor associated protein 1(TRAP1)
GBP1	guanylate binding protein 1(GBP1)
DDX58	DExD/H-box helicase 58(DDX58)
DECR1	2,4-dienoyl-CoA reductase 1, mitochondrial(DECR1)
FARP1	FERM, ARH/RhoGEF and pleckstrin domain protein 1(FARP1)
GAPDH	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)
PPIA	peptidylprolyl isomerase A(PPIA)
PPIB	peptidylprolyl isomerase B(PPIB)
RPL30	ribosomal protein L30(RPL30)

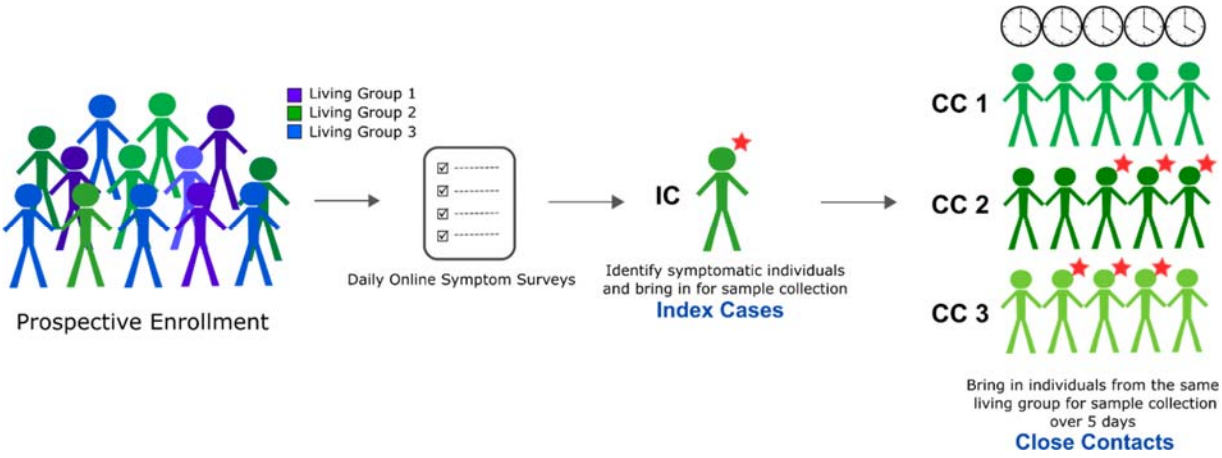

Figure s1. Students were enrolled at the beginning of the school year. Index Cases (ICs) were identified when daily symptom surveys indicated potential respiratory illness. This flagged serial sampling of their Close Contacts (CCs), some of whom would also develop RTI during the 5-day observation window (red stars). This unique model permits blood sampling of some Close Contacts during early and presymptomatic phases of illness.

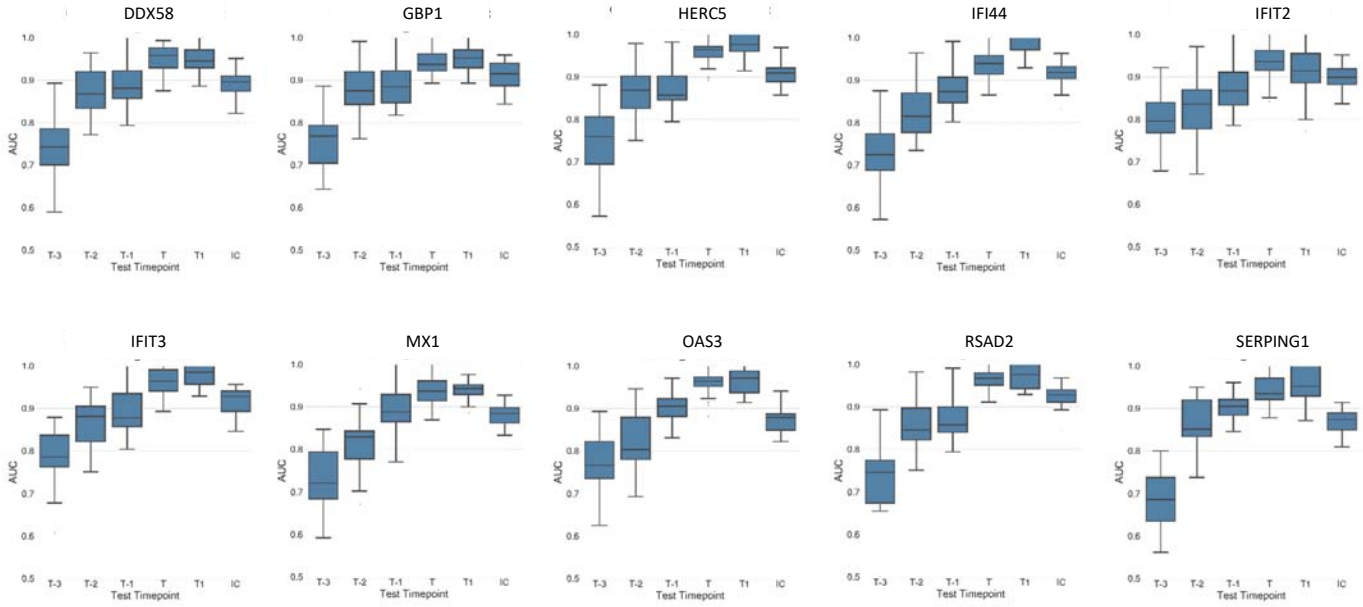

Figure s2. Diagnostic performance of individual genes at each timepoint. We evaluate the diagnostic performance of individual genes by calculating auROC at each test time point from looking at individual gene expression values from the subjects in each model instance against the outcome (Symptomatic Shedders vs Asymptomatic Non-Shedders). That is, we directly use the gene values instead of a predicted probability; this procedure is equivalent to generating an ROC curve by thresholding gene values and calculating the AUC of this curve. The performance of the top 10 individual genes at discriminating symptomatic viral infection (selected by calculating mean AUC across all model instances at all test time points) are shown above.

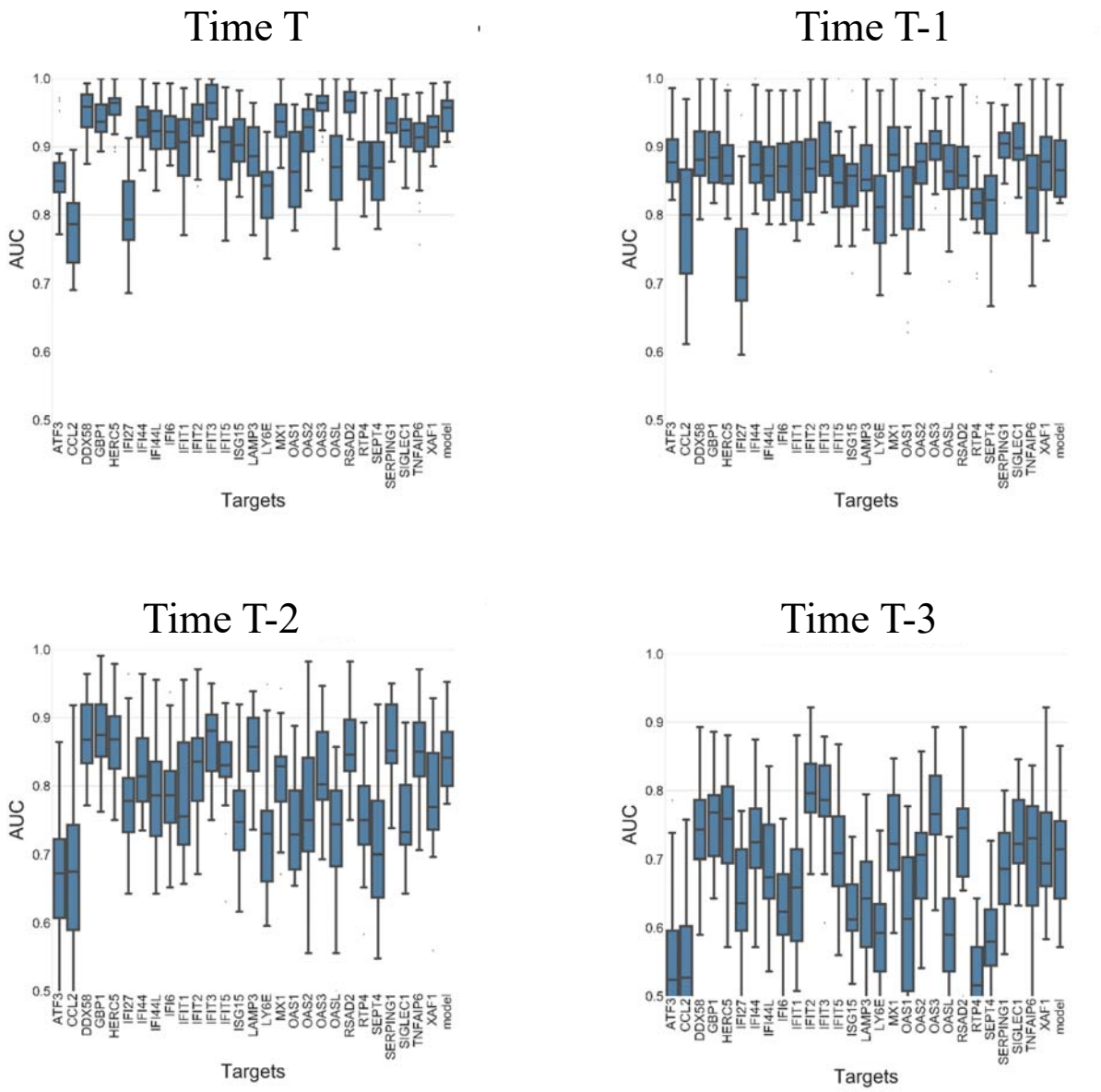

Figure s3. Performance of individual genes across time. Classification performance of each individual gene compared to the full 36-gene model (right-most model in each panel). Each panel represents the performance of each individual gene in the model (listed on the x-axis) at each indicated timepoint.

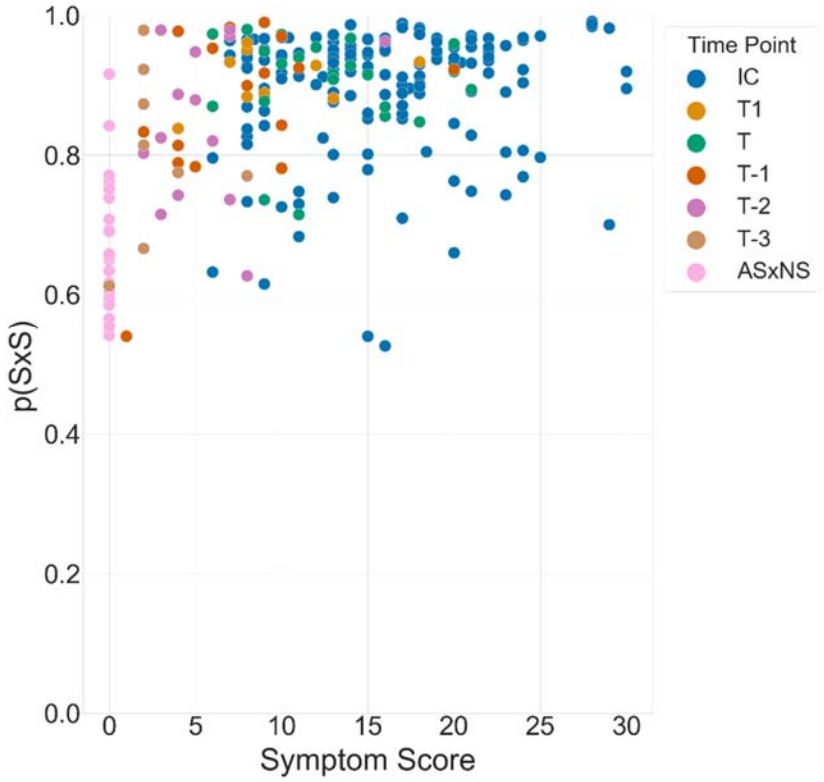

Figure s4. Disease severity vs Signature strength. Symptom severity varies by timepoint, gradually increasing over time, but no correlation is seen between strength of the transcriptomic signature (relative probability of being a symptomatic shedder, p(SxS)) and symptom severity at any timepoint.

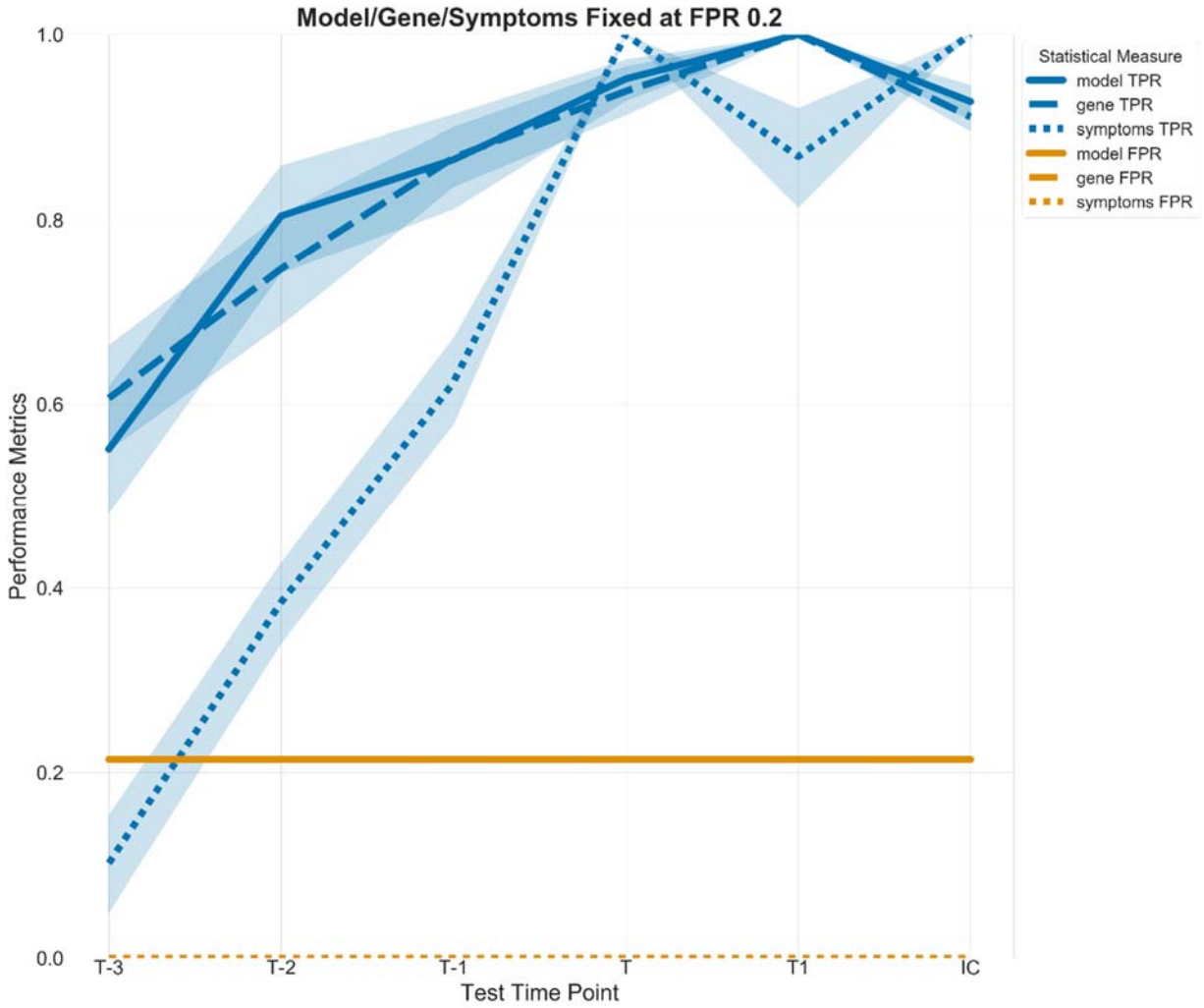

Figure s5. Signature performance vs symptomatic diagnosis. The transcriptomic model outperforms symptoms alone at all times except time T (where by definition all subjects met symptom criteria). Here we demonstrate the TPR at the different test timepoints for the model and the best performing gene IFIT3 ("gene TPR", see Figure S2) when the FPR is fixed at 0.2, and that of the symptom score when it is thresholded at 6. As expected, later timepoints and timepoint 'IC' have higher TPRs. Note that by definition, since we define symptomatic patients as those with a symptom score larger than 6, the FPR for the symptom scores will be zero and the TPR for symptoms at time 'T' and 'IC' will be 1.

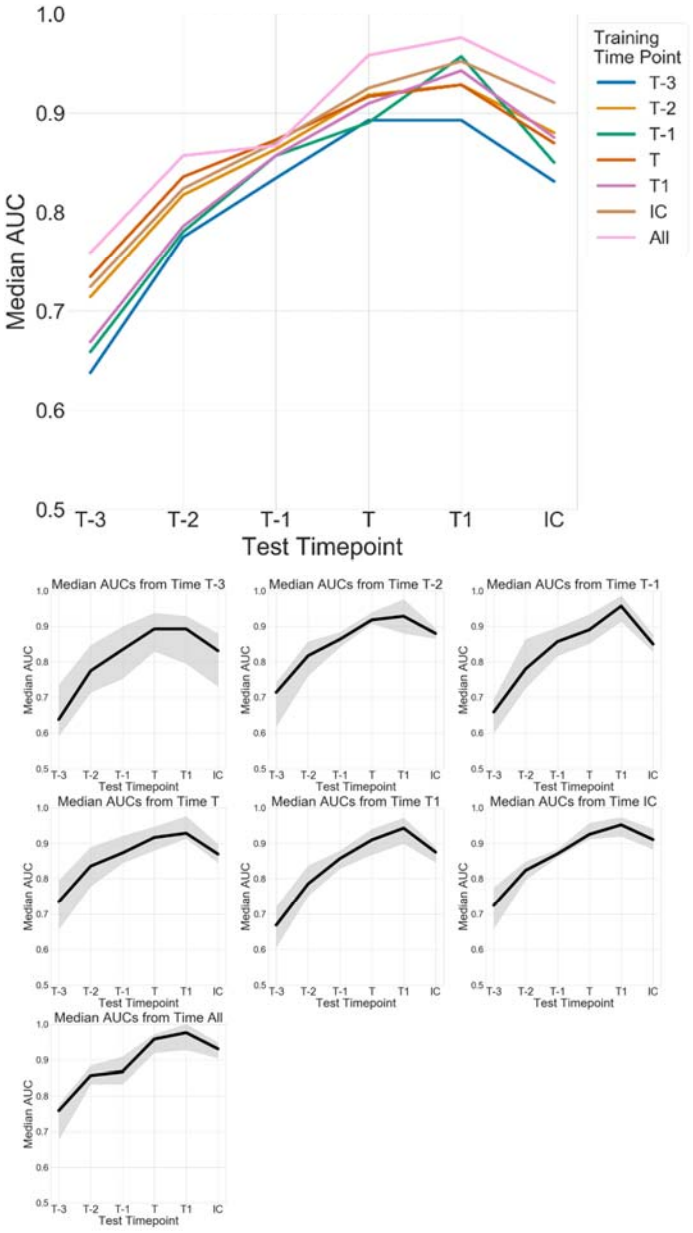

Figure s6. Effect of training timepoint on signature performance. Median signature performance variation by training time point. Training the model specifically on early times offers no significant improvement in signature performance. For the primary analysis (Figure 3) subjects in each model instance can be assigned to any time point in training. Here, we present results when training on a specific single timepoint instead. Depicted is the median auROC across all model instances for each test time point. The auROC increases over time and is also high at time 'IC' no matter which training timepoint is used. However, we also see that training on early timepoints does not offer any significant performance benefit.

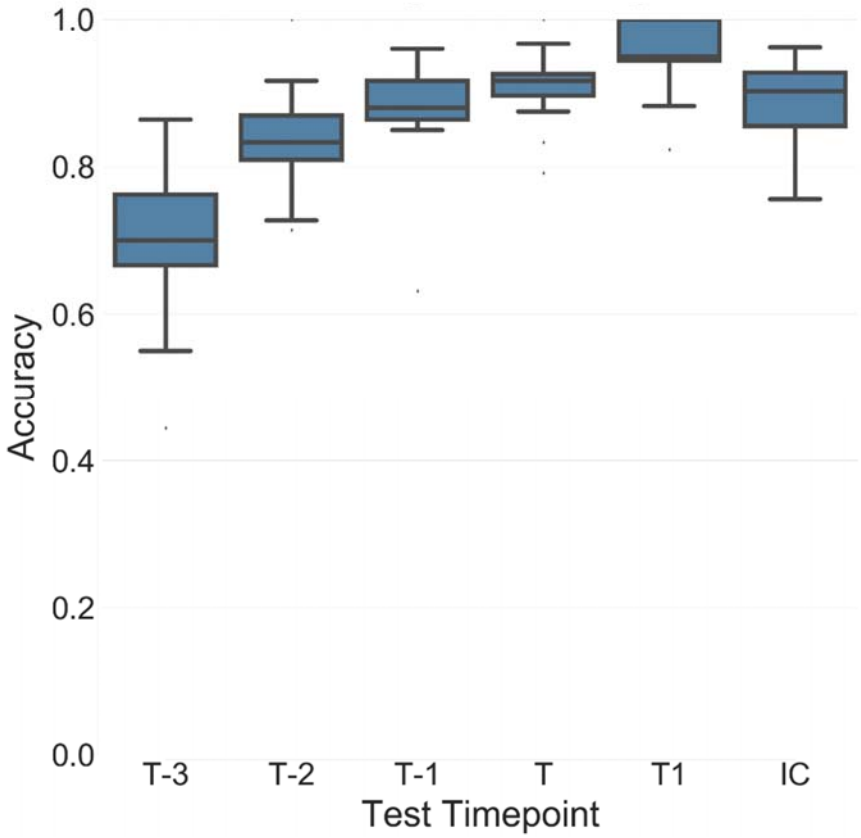

Figure s7. Signature accuracy by timepoint, across all viruses. For each of the 25 model instances, we have as outputs a set of predicted probabilities for the subjects at each test timepoint; we use these predicted probabilities to generate ROC curves for each test timepoint. For each ROC curve, we compute and report the accuracy corresponding to the sensitivity and specificity at which the Youden-J statistic is maximized.